온톨로지 기반 대학정보 검색 시스템의 설계 및 구현

박종훈* · 김철원**

Implementation and Design of College Information Retrieval System Based On Ontology

Jong-hoon Park* · Chul-Won Kim**

요 약

오늘날 효과적인 정보검색을 위해 지능형 검색에 대한 다양한 기법들을 사용하고 있다. 이 중에서 효과적인 검색 방법은 온톨로지 기술을 적용하는 것이다. 온톨로지는 시맨틱웹에서의 핵심기술이라 할 수 있다. 시맨틱웹에서 온톨로지 기술은 간단하면서 정확하게 추론엔진을 통하여 관련 정보를 검색하는데 사용될 수 있다. 본 논문에서는 대학, 대학원, 구성원을 중심으로 정보를 검색할 수 있는 온톨로지 기반 대학정보검색 시스템을 설계 및 구현을 하고자 한다. 대학, 대학원, 구성원 정보들의 계층구조를 수집하였으며, 온톨로지 개발도구인 *protege* 에디터를 이용하여 대학정보를 온톨로지로 설계하기 위해 설계된 대학정보 온톨로지를 *protege* 에디터의 추론기능을 이용하여 검증하였으며, 검증된 온톨로지는 지나 추론엔진을 적용하여 웹서비스 할 수 있도록 대학정보검색 시스템을 구현하였다.

ABSTRACT

Currently, in order to develop an intelligent search engine to help users retrieve information effectively, many metodes have been used. The effective retrieval methods of these methods use ontology technology. Ontology technology is the core of the Semantic Web. In the Semantic Web, ontology technology can be used to retrieve related information through the inference engine more accurately and simply on the Semantic Web. In this paper, we implement and design college information retrieval based on ontology to retrieve college class, graduate school class and person class. We have collected the hierarchy structure about the College, graduate school and person informations, and we have used *protege* editor of the ontology developing tool to design some ontologies with the College informations collected. We also tested the designed ontology with the Inference Engine(*Pellet*) of *protege* editor, and implemented college information retrieval system using Inference Engine(*Jena*) for web services.

키워드

온톨로지, 시맨틱 웹, 추론엔진, OWL

Key word

Ontology, Semantic Web, Inference engine, OWL

* 정회원 : 중부대학교 컴퓨터학과(주저자)

** 종신회원 : 호남대학교 컴퓨터공학과(교신저자, cwkim@honam.ac.kr)

접수일자 : 2011. 11. 01

심사완료일자 : 2011. 12. 21

I. 서 론

웹 이용이 보급화되면서, 사용자는 늘어나고 공유하는 정보의 양이 예상하지 못 할 정도로 증가하고있다. 또한 검증되지 않는 정보의 폭발적인 증가로 인해 사용자가 원하는 정보를 찾는 것이 어려워지고 있다. 그리고 현재의 웹은 문서들 간의 단순 연결을 통해서 정보를 제공하며, 이러한 연결을 통해서 상호 관련된 정보를 쉽게 찾을 수 있다는 장점이 있지만, 수집 및 검색된 정보가 사람에 의해서 해석되고 정제되어야 한다는 문제점이 있다 [1][2].

시맨틱 웹은 정보의 의미를 개념으로 정의하고 개념들 간의 관계성을 표현함으로써, 문서들 간의 단순 연결이 아닌 의미 연결을 통해서 보다 정확하고 효율적인 정보 검색을 가능하게 한다. 현재 시맨틱 웹의 핵심 기술인 온톨로지(Ontology)구축에 대한 연구가 활발히 진행되고 있다. 온톨로지는 의미의 기본단위인 개념과 그들 사이의 관계들로 표현되는데, 이렇게 표현화된 의미 단위와 표현 방법을 사용함으로써 사람과 컴퓨터, 컴퓨터와 컴퓨터가 의미모호성 없이 정보를 해석할 수 있다. 온톨로지는 스스로 개념화 과정과 추론 과정을 통해 사람이 원하는 정보를 정확하게 찾아줄 수 있다 [3][4]. 현재 대학 홈페이지는 점차적으로 복잡화되고 있으며, 대학 정보 내용과 검색단어가 점차적으로 증가하고 있다. 따라서 대학정보의 검색의 효율성을 증대시키고자 온톨로지를 활용한 대학정보 검색에 대한 기법연구가 필요하다.

따라서 이에 대한 필요성으로 본 논문에서는 시맨틱 웹의 핵심 기술인 온톨로지를 이용하여 온톨로지 기반의 대학 정보 검색 시스템을 설계 및 구현하였다.

II. 온톨로지 관련기술

2.1. 온톨로지

2.1.1. 온톨로지의정의

온톨로지는 단어와 관계들로 구성된 일종의 사전으로서 생각할 수 있으며, 그 속에는 특정 도메인에 관련된 단어들이 계층적으로 표현되어 있고, 추가적으로 이를 확장할 수 있는 추론 규칙이 포함되어 있어, 웹 기반의 지식 처리나 응용 프로그램 사이의 지식 공유, 재사용 등이 가능토록 되어 있다 [2].

2.1.2. 온톨로지의 구성

일반적으로 온톨로지는 클래스(class), 인스턴스(instance), 관계(relation), 속성(property), 네가지 구성원소로 구성되어 있다.

2.2. OWL(Web Ontology Language)

2.2.1. OWL의 정의와 종류

웹의 리소스를 기술하는 온톨로지 언어로서 2004년에 W3C에 의해서 권고된 것이 OWL이다. OWL에는, 기술논리(Description Logic)를 토대로 하는 OWL DL과 이에 대한 단순화된 버전으로서 OWL Lite, 그리고 같은 온톨로지 언어 RDF스키마의 확장으로서 보다 자유로운 기술이 가능한 OWL Full과 같은 3가지 언어가 있다 [5].

2.2.2. OWL의 클래스 및 클래스의 논리조합

클래스는 같은 성질을 갖는 리소스들을 그룹화하고 공통성질을 논리적으로 표현하기 위한 기능을 제공한다. 클래스에 의한 개념의 체계적인 기술은 온톨로지의 핵심이 된다. OWL에서는 기본 어휘 owl:Class에 의해서 클래스를 표현한다. OWL클래스는 URI참조(Uniform Resource Identifiers)를 이용하여 명명된다. 클래스의 인스턴스는 개체(Individual), 인스턴스의 집합은 클래스의 외연(Class Extension) 이라고 부른다.

2.2.3. OWL의 속성

OWL의 속성에는 개체들 간의 관계를 기술하는 개체값형 속성(individualvalued Property = owl:Object Property)과, 개체를 데이터값과 관련시키는 데이터값형 속성(datavaluedProperty = owl:DatatypePropetry)이 있다. 또한 온톨로지의 기술에 사용되는 특수한 타입으로 주석형 속성과 온톨로지 속성이 제공되고 있다.

2.3. Protege 편집기

Protege 편집기는 스텐포드대학교에서 개발된 자바 기반의 온톨로지 에디터이다. 보다 정교한 어휘를 작성하기 위해 OWL을 이용하여 온톨로지를 구축한다. OWL의 RDF/XML은 그래프 도구 및 텍스트 편집기 환경에서는 복잡하므로, Protege와 같은 온톨로지 편집기가 편리하다[6]. Protege를 사용한 온톨로지 설계절차는 다음과 같다.

1) 클래스 계층구조 만들기

- 2) 속성 작성하기
- 3) 클래스의 속성 제약 정의하기
- 4) 온톨로지의 주석과 연동
- 5) 인스턴스의 작성

온톨로지 에디터인 **Protege** 편집기 내에는 온톨로지 검증을 위한 도구를 제공하고 있다. **Protege** 에디터 메뉴에 "Reasoner"에서의 **Pellet**를 사용하여 기술내용의 일관성(consistency)과 다른 텍스트 박스에 입력한 그래프가 해당 온톨로지로부터 논리적으로 잘 구성되는지 등을 검증해 준다. 또한 **Protege** 에디터의 "DL Query" 탭에서 **The Manchester OWL Syntax** 구문을 사용하여 온톨로지 검색도 가능하다. 이것은 클래스(Class), 속성(Property), 개체(Individuals) 등에 대해서 다양한 방식으로 온톨로지 검색을 할 수 있다.

2.4. Jena 추론엔진

Jena는 시맨틱 웹 응용을 만들기 위해 HP 연구소에 의해서 만들어진 자바 기반의 프레임워크이다. **Jena**는 **RDF**, **RDFS**, **OWL**, **SPARQL** 그리고 **Rule** 기반 추론 엔진을 위1한 프로그램 환경을 제공한다. 즉 온톨로지 모델과 개발자 사이에서 지식 관리 역할을 하는 계층인 것이다 [7].

Jena는 자바로 구현되어 있어서 자바언어가 갖는 모든 장점을 가지고 있으며, 기본적인 **RDF** 파서도 제공하며 내부적인 추론은 그래프 매칭 방법을 이용하여 접근한다. 지나의 프레임워크는 다음과 같다.

- 1) **RDF API**: **RDF**의 기본 구조인 트리플 구조를 모델링한 **Statement** 클래스나 **Model**, **Subject**, **Object**, **Property** 각각을 독립된 클래스로 제공하여 **RDF** 모델을 직관적으로 생성 및 변경하는 코딩을 할 수 있다.
- 2) **Reading and writing RDF in RDF/XML, N3 and N-Triples**: **RDF** 모델을 **RDF/XML**이나 **N3** 혹은 **N-Triples** 형태로 읽거나 쓰는 방법을 제공한다.
- 3) **OWL API**: 현재 시맨틱 웹 추론 엔진 부분에서 가장 활발한 업데이트와 다양한 기능을 제공하고 있다. **OWL**와 관련된 **API**를 제공하며 검색, 생성, 삭제, 편집 등 관련 작업을 진행할 수 있다.
- 4) **RDQL(RDF Data Query Language)**: **RDQL**은 **RDF** 모델 상에서 조건에 맞는 **Subject**, **Object** 혹은 **Property** 질의를 통해 작업을 한다.

III. 온톨로지 기반 대학정보 검색 시스템의 설계 및 구현

3.1. 온톨로지 설계

온톨로지 기반의 학교정보 구축을 설계하기 위하여, 본 논문에서는 특정한 한 학교의 관련 정보에 대해서, 개발 도구 **Protege** 편집기를 이용해 온톨로지를 설계한다. 전체 설계 절차는 대학정보수집, 클래스 분류 및 속성분류, 인스턴스 생성, **OWL** 생성과 **OWL** 생성에 대한 검증 단계로 설계하였다.

3.1.1. 클래스(Class) 모델 설계

학교 온톨로지 구축하기 위하여 본 논문에서는 특정한 대학교를 선택하여, 대학 정보를 수집하여, 온톨로지를 작성하도록 한다. 전체 정보는 너무 방대하여 일부 영역만을 제한하여 설계하였다.

각각 클래스간의 트리형 클래스 계층구조는 다음의 그림 1과 같다.

그림 1. 클래스 계층구조
Fig. 1 Hierarchy of classes

3.1.2. 개체(Individual)의 설계

본 논문에서는 대학 클래스와 대학원 클래스에 속하는 각 학과를 나타내는 서브 클래스들에 대해서 각 학과의 개설과목들을 개체로 생성하였다. 사람 클래스의 서

브클래스인 교수님 클래스, 행정직원 클래스와 학생 클래스에 대해서 각각 구성원들의 이름을 개체로 생성하였다.

3.1.3. 속성(Property)의 설계

본 논문에서 각 클래스 및 개체들에 대해서 다음과 같은 주요 속성을 부여한다. **teaches** 속성은 실제로 존재하는 교수 및 담당과목 간의 관계를 나타내기 위한 부여된 속성이다. 교수는 특정 담당과목을 강의한다는 뜻을 의미한다. **isTaughtBy** 속성은 담당과목과 교수 간의 관계를 나타내기 위한 부여된 속성이다. 과목은 교수의 강의를 통해서 실시한다는 뜻을 의미한다. **takes** 속성은 학생과 수강과목 간의 관계를 나타내기 위한 부여된 속성이며, 학생들이 과목을 수강한다는 뜻을 의미한다. **workat** 속성은 교직원과 소속 부서와의 관계를 나타낸다. 예로 교수가 어느 학과에 소속되어 있는지를 결정할 때 사용한다. **majorin** 속성은 학생과 학과와의 관계를 나타낸다. 즉 어느 학생이 어느 학과에 소속되어 있는지를 알 수 있는 속성이다.

3.1.4. 온톨로지의 추론 및 검증

Protege 편집기에 외부 온톨로지 추론(Pellet)을 위한 도구들을 제공하고 있다. 이를 적용하여 대학정보 온톨로지인 OWL 파일이 정상적으로 만들어 졌는지를 확인한다. 추론기를 통하여 전체 클래스 관계가 정상적으로 구축되었는지를 추론한 후 클래스 구성과 이를 통한 검색을 하여 확인하였다.

그림 2. 검색어 “컴퓨터학과”에 대한 Individuals 검색의 예

Fig. 2 Retrieval example by Individuals for “컴퓨터학과” word

그림 2는 Protege 편집기 DL Query 탭에서 “컴퓨터학과”를 검색하는 화면이며 옵션선택에서 Individuals 검색을 선택한 예이다.

3.2. 온톨로지 기반 대학정보검색 시스템 구현

본 논문에서 설계된 대학정보 온톨로지를 웹서비스 환경에서 적용하기 위해 다음 표 1과 같은 환경을 구현하였다.

표 1. 시스템 구현환경
Table. 1 Implementation environment of system

웹 서버 환경	웹 서버	Apache Tomcat 6.0.33
	웹 프로그램	JSP
온톨로지 구축	언어	OWL
	개발 도구	Protege 4.1_beta
온톨로지 검색	추론엔진	Jena
	인터페이스	Java 1.6.0_29

다음 그림 3은 구현된 대학정보검색 시스템의 내부 처리 절차를 보여주고 있다. 사용자인터페이스에서 사용자의 검색단어를 입력하면 OWL 추론 엔진이 단어를 분류하여 클래스인지 개체인지를 식별한다. 온톨로지 저장소에는 설계된 대학정보 OWL 파일이 구성되어 있으며, 검색 단어와 관련된 클래스, 개체와 속성들을 검색하여 결과를 사용자 인터페이스 화면에 출력한다.

그림 3. 대학정보 검색 시스템 처리 절차

Fig. 3 Process sequences of college information retrieval system

3.2.1. 구현 결과

다음 그림 4는 사용자가 검색단어를 “컴퓨터학과”를 래스를 입력한 경우의 입력 화면이다. 대분류에서 컴퓨터학과는 대학/학과에 속하므로 사용자가 선택한 후 검색단어를 입력한다.

3.3. 고찰

본 논문에서는 온톨로지에 기반한 대학정보검색을 위하여 기본적인 분야의 클래스를 모델화하여 설계 및 구현 하였다. 그러나 실제로는 클래스와 속성 등에 대한 종류가 더욱 세분화되고 다양화할 필요가 있다. 속성 분류에 대해서 더 깊이 연구할 필요가 있으며, 속성부여의 단계는 여러 가지 상황을 고려하여 정하여야 한다. 본 논문에서는 기본적인 속성에 대해서만 부여하였으나 향후 더 다양한 속성이 필요하다.

추론엔진과 결합해서 효율적인 검색 시스템을 구현할 때 대학정보 온톨로지를 어떻게 설계하느냐에 따라 검색 정확도와 사용자의 불편을 최소화할 수 있다. 또한 대학정보 온톨로지에 대한 데이터양이 미약하므로 향후 데이터양을 추가 보완하여 검증할 필요가 있다. 검색 단어 입력시에 사용자는 온톨로지 클래스에 대한 약간의 지식이 필요하며, 제한된 입력만 되어 있으므로 향후 이러한 문제점을 개선할 필요가 있다.

IV. 결론

현재 우리는 정보 홍수의 시대에 살고 있으며, 인터넷 공유 정보의 폭발적인 증가로 효율적인 검색의 필요성이 제기되고 있다. 더불어 학교 정보에 대해서도 기존의 단순 키워드 검색 방식은 사용자들에게 원하는 정보를 제공하는데는 한계가 있다. 이에 대한 대안으로 온톨로지 검색 기술을 적용하여 사용자에게 의미검색과 편리성과 검색 효율성을 높이기 위한 시스템이 필요하다.

본 논문에서는 이러한 문제점을 해결하고자 학교 정보 검색에 대해서 효율적인 검색 방법인 온톨로지기술을 적용하였다. 대학, 대학원, 구성원을 중심으로 정보를 검색할 수 있는 온톨로지 기반 대학정보검색 시스템을 설계 및 구현하였다.

현재 온톨로지의 개발에 대해서 W3C에 비롯한 많은 학자나 연구진이 활발히 진행하고 있다. 온톨로지 검색은 일반 검색에 비해 정확하고 효율적인 많은 장점을 갖고 있기 때문에 앞으로 다양한 분야에 보급화되어 정보의 획득에 주요한 이용 수단으로 될 전망이다.

참고문헌

- [1] 박지현, "온톨로지에서 RDF를 이용하여 의미 정보를 확장하는 방법에 관한 연구" 이화여 자 대학교 대학원, 석사학위 논문 2005, 7.
- [2] 神奇 正英 "시맨틱 웹을 위한 RDF/OWL 입문", 황석형/양해술 공역, 홍릉 과학 출판사, 2008년.
- [3] 조성환, "관계형 데이터 베이스에 기반한 효율적인 OWL 저장 관리 시스템 설계 및 구현", 한국 교원대학교 대학원, 석사학위 논문, 2007, 2.
- [4] 김수경, 안기홍, "시맨틱 웹 응용을 위한 웹 온톨로지 구축기법", 정보처리학회논문지D, 제15권 1호, 2008년2월
- [5] Deborah McGuinness, Frank van Harmelen, "OWL Web Ontology Language Overview", W3C Recommendation, 2004.
- [6] Protege ontology editor and knowledge acquisition system, <http://protege.stanford.edu>
- [7] Jena - A semantic web Framework for java, <http://jena.sourceforge.net>

저자소개

박종훈(Jong Hoon Park)

1995년 광운대학교 전자계산기
공학과(공학박사)
1995년~1998년 한국전산원
선임연구원

1999년- 현재 중부대학교 컴퓨터학과 교수

※관심분야: XML 웹서비스, XML 검색, 시맨틱 웹

김철원(Chul Won Kim)

1996년 광운대학교 컴퓨터공학과
(공학박사)
1988년~현재 호남대학교
컴퓨터공학과 교수

※관심분야: XML 웹 서비스, 이미지 처리 및 검색